STATISTICAL ADAPTATION ALGORITHM IN CONTINUOUS SHORT-RANGE RUNOFF FORECASTS

V. I. Koren

Meteorologiya i Gidrologiya,
No. 3, pp. 80-84, 1984

UDC 556.16.06

Proposed is an algorithm which increases the accuracy of short-range forecasts of discharge and levels of water by considering initial conditions. It is assumed that errors in the input data are random. The algorithm is checked for a large number of cases.

In making continuous short-range forecasts of discharge new data on the predicted element is usually supplied. This data characterizes the state of the discussed water object at the moment of the forecast (initial conditions). Naturally, the effectiveness of the method used for forecasting depends in many ways on the method of calculating the initial conditions.

In the practice of hydrological forecasting numerical methods are often used, which clearly do not include initial conditions, making their implementation difficult. These methods may include the plan for forecasting water discharge based on the use of lag curves. In this case the expression for determining the predicted value may be written in the following form:

$$\widehat{Q}(t_0+\Delta)=\sum_{i=1}^{m}k_i\sum_{j=t_i}^{t_0+\Delta}q(i,\ j)P(i,\ t_0+\Delta-j+1), \tag{1}$$

where $\widehat{Q}(t_0+\Delta)$ is the value predicted at the output with the forecast term of Δ units of time, t_0 is the moment of the forecast, m is the number of inputs for which the lag curves are assigned, t_i is the beginning of the calculation, i.e., the moment of time at which deformation for the i-th input begins to be considered ($t_i \leq t_0$), q(i, j) are the ordinates of the i-th input at the j-th moment of time, P(i, j) are the ordinates of the lag curves, and k_i are the discharge coefficients, determined from the conditions of preservation of volumes.

Data at inputs may represent in [1] either the measured water discharge at corresponding sites or the computed values of water yield from the area related to the discussed input. Lag curves and discharge coefficients are usually determined according to archival data and are preserved unchanged in making the forecasts. The beginning of the calculation depends on the form of the lag curves and is selected with the calculation that the difference between the moment of forecast and the beginning of calculation is not less than the "memory" (number of significant ordinates) of the lag curves (δ_p). The model used (1) is not completely adequate for the actual process, and the input data is assigned with erros (especially if water yield is used rather than actual water discharge). This may lead to the fact that the calculated water discharge $\widehat{Q}(t)$ for $t_i<t<t_0$ will differ from that already available at the moment of the forecast of actual values of Q(t). The latter circumstance leads to a decrease in the accuracy of the forecasts, especially with small forecast term, since the process described by dependence (1) has inertia (higher with higher δ_p). The period during which accuracy decreases due to inaccuracy of calculation up to the moment t_0 depends on the form of the lag curves, which indirectly is determined by the value of δ_p.

To exclude possible accumulation of errors at the moment of forecast in this plan we must consider the initial conditions. In general form the relationship considering the initial conditions can be represented as follows:

$$\tilde{Q}(t_0+\Delta)=\hat{Q}(t_0+\Delta)+\Phi(Q,\Delta), \qquad (2)$$

where $\Phi(Q,\Delta)$ is some operator which assimilates the factual data at the output, which are supplied at the moment of forecast.

For some partial cases we obtained analytical expressions. Thus, in approximation of the lag curve by gamma distribution one of whose parameters (n) takes only whole values, we obtained the following relationship [2]:

$$\Phi(Q,\Delta)=\sum_{i=1}^{n}C_i\frac{1}{(i-1)!}\left(\frac{\Delta}{\tau}\right)^{i-1}e^{-\frac{\Delta}{\tau}}, \qquad (3)$$

where n, τ are the parameters of the lag curve, and C_i are constants of integration which must be determined from initial conditions.

If n = 1, the coefficient C_1 is equal to the actual water discharge on the day of the forecast $Q(t_0)$, since where $t_i=t_0$ $\hat{Q}(t_0)=0$. In this case the forecast will always begin with the actual discharge, and the error of the input data up to the moment t_0 will not affect the forecast results. Methods of determining the coefficients C_i where $n \neq 1$ are presented in [5]. We note only that the greater the value of n, the lower the effectiveness of operator (3), since with the increase in n there is a strong increase in the effect of the inaccuracy of the input data on the process of determining the coefficients. Due to this a method for decreasing the parameter n without substantial loss of accuracy of transformation was proposed in [6].

Drawbacks of this method of correcting the forecasts are obvious: the approximation of the lag curve is firmly assigned, the parameter n has strong limitations, and errors of input data are clearly not considered.

Proposed in [1,2] were methods of correction based on the replacement of the actual input hydrograph by an arbitrary one, which decreases the errors of the calculation up to the moment of the forecast. In one case [1] only one ordinate is changed with such a calculation so that the calculated water discharge at the output on the day of the forecast coincided with the actual discharge; in another [2] all ordinates on the interval (t_1, t_0) are changed. Thus, in the first method an excessively large role is given to the discharge on the day of the forecast, and in the second, on the other hand, this discharge may not have a marked effect on the correction, which thus lowers its effectiveness. Moreover, the random nature of the errors is not considered in these approaches.

Considering the errors of the input data the relationship (2) may be represented in the following form:

$$\tilde{Q}(t)=\hat{Q}(t)+\sum_{i=1}^{m}k_i\omega_i\sum_{j=t_1}^{t}\eta(j)P(i,\ t-j+1), \qquad (4)$$

where $\eta(j)$ is the error at the input at the j-th moment of time, and ω_i is the percentage of runoff of the i-th input (in terms of runoff in a closed site it was usually determined as the ratio of volumes of runoff during genetically homogeneous periods).

To use relationships (4) in forecasts we must somehow assign the vector η, since we do not actually know the errors of input data.

Let us assume that the errors follow a normal law of distribution with center $\bar{\eta}$ and dispersion D_η. We model the vector η_0 with parameters $\bar{\eta}=0$ and D_n = ' (white noise) in the time interval $(t_1,\ t_0+\Delta_m)$ (Δ_m is the maximal forecast term). In each actual case the forecast of the parameters of distribution of errors are unknown and may differ from those used in modeling the vector η_0. To estimate them we construct an algorithm based on minimizing the following functional:

$$R(\eta_0, \bar{\eta}, D\eta) = \sum_{j=t_p}^{t_0} \alpha_j [Q(j) - \tilde{Q}(j)]^2, \tag{5}$$

where $\bar{\eta}$, D_η are the sought parameters of distribution of errors, $\tilde{Q}(j)$ is water discharge, calculated according to (4), $Q(j)$ is the actual water discharge, α_j are the coefficients which permit us to assign different weight to the deviations of the calculated discharge from the actual values at different moments of time, and t_p is the moment of time from which calculation errors are considered.

In calculating $\tilde{Q}(t)$ according to (4) the vectors of errors is recalculated with consideration of the parameters $\overline{\eta_k}$ and $D\eta_k$, which are obtained in the process of minimizing functional (5):

$$\eta_k = \eta_0 D\eta_k + \overline{\eta_k}. \tag{6}$$

To minimize functional (5) we can use the methods of optimization and obtain estimates of $\overline{\eta^*}$ and D^*_η which correspond to the vector η_0 with some value of $R^*(\eta_0, \overline{\eta^*}, D^*\eta)$. Since the vector of errors is generated for a brief time period, it may not be optimal for the given input data. Because of this the procedure of minimizing functional (5) must be repeated p times, modeling each time the new vector η_0 with the parameters $\bar{\eta} = 0$ and $D_\eta = 1$. As a result we obtain a series of estimates of $(\overline{\eta_i}, D^*\eta_i)$ and values of R^*_i which correspond to the original vectors η_{0i} $(i=1, 2, \ldots, p)$.

For use in the forecast we select the estimates of the errors and the original vector which correspond to the least value of R^*_i from the obtained series. Such an adaptation of the plan to the input data should be repeated for each case of making a forecast. Therefore it assumes that data on the predicted value enters and is assimilated continuously.

In the practical implementation of the discussed algorithm we need to assign time intervals $\Delta t_i = t_0 - t_i + 1$ and $\Delta t_p = t_0 - t_p + 1$, and also weighted coefficients α_j. Since t_p should be greater than or equal to $t_1 + \delta_p$, it is worthwhile to posit $\Delta t_i = \Delta t_p + \delta_p$. It is considerably more complicated to select a priori the values of Δt_p and α_j. To decrease the definite nature in selecting weighted coefficients we can use an indicative dependence

$$\alpha_j = \alpha^{\Delta t_p - j}, \tag{7}$$

where α is the coefficient of the decrease in the effect of calculation errors from t_0 to t_p, and $0 < \alpha < 1$.

The discussed algorithm was used to correct short-range forecasts of the discharges and levels of water in basins of the Pechora, Onon, and Neya Rivers (more than ten sites were observed). For the first two drainage basins the forecast was based on transformation of hydrographs assigned at input sites. For sites where observations were made only on water levels, arbitrary curves of discharge were constructed by the method proposed in [3]. For the Neya R. predictions were made on the basis of the model of formation of thaw and rain runoff [4]. In this case the input data in (1) consisted of calculated values of water yield. Verified forecasts were compiled daily in the period of the open river channel according to data for 7-15 years. The forecast term for different sites varied from two to eight days. As input data in the period of the forecast term we used water discharge predicted for the subjacent area. For individual areas (drainage basins) the input data for the period of the forecast term were extrapolated. In general for each site we compiled up to 5000 forecasts of different terms.

Calculations showed that the proposed adaptation algorithm significantly increases forecast accuracy. The greatest effect is achieved, as should be expected, for predictions with short forecast terms (1-2 days). For all the discussed sites the forecasts for days without consideration of initial conditions turn out to be practically ineffective $(s/\sigma_\Delta > 1)$, whereas when initial conditions were considered, the value of s/σ_Δ was always less than one. With increase in the forecast term the effectiveness of correction decreases and estimates of the forecasts gradually approach the estimates obtained without considering initial conditions. Table 1 presents the mean estimates of the forecasts of s/σ_Δ over 10-15 years for different forecast terms Δ, expressed in days, at some sites.

Table 1

Mean Estimates of Forecasts of s/σ_{Δ} for Some Sites Obtained with (Numerator) and without (Denominator) the Use of the Adaptation Algorithm

River — point	Forecast term, days							
	1	2	3	4	5	6	7	8
Onon R. — village of Bytev	0,35/0,52	0,40/0,42	0,53/0,54	0,69/0,69				
Onon R. — village of Chindat-1	0,36/1,03	0,34/0,60	0,34/0,47	0,37/0,43	0,64/0,69	0,61/0,62	0,68/0,68	
Onon R. — Olovyannaya station	0,36/1,04	0,36/0,66	0,36/0,50	0,36/0,42	0,38/0,41	0,39/0,42	0,44/0,49	0,57/0,57
Onon R. — village of Chiron	0,51/0,94	0,46/0,65	0,47/0,58	0,47/0,54	0,49/0,53	0,52/0,54	0,57/0,58	0,60/0,65
Naya R. — town of Busleevo	0,60/1,14	0,70/0,90	0,67/0,72	0,66/0,66				

In processing the adaptation algorithm we studied the effect of the parameters α and Δt_p on forecast accuracy. It turned out that these parameters have greatest effect on the accuracy of forecasts with small terms. With a decrease in the parameter Δt_p the role of α decreases greatly, and where $\Delta t_p = 2$ its value has practically no effect on the results. For most cases the best accuracy was achieved where $\Delta t_p = 3$ and $\alpha = 0.3$-0.5. With the use of the proposed algorithm for other objects it is worthwhile to select beforehand the optimal value of the parameter Δt_p (where $\alpha = 0.3$-0.5), making verification calculations according to archival data for values of Δt_p from 1 to 6 days.

It should also be noted that in a striking number of cases the extension of the vector η with the same characteristics to the period of the forecast term led to a decrease in forecast accuracy. Therefore, in all later calculations the vector η was modeled only up to the moment of the forecast (t_0), and it was assumed equal to zero in the forecast term.

REFERENCES

1. A. P. Zhidikov et al., Methods of Calculating and Predicting High Water for a Cascade of Reservoirs and River Systems [in Russian], Gidrometeoizdat, Leningrad, 1977.
2. V. I. Koren and L. S. Kuchment, "Mathematical model of the formation of flash floods, optimization of its parameters, and use in hydrological forecasts," Meteorologiya i Gidrologiya, no. 1, 1971.
3. V. I. Koren, "Determining the parameters of lag curves in the absence of observations of water discharge in a closed site," Trudy Gidromettsentra SSSR, no. 207, 1978.
4. V. I. Koren, "Modeling processes of formation of river runoff of a forest zone of the European territory of the USSR," Meteorologiya i Gidrologiya [Soviet Meteorology and Hydrology], no. 10, 1980.
5. V. I. Koren, "Mathematical model of flash and thawed discharge in hydrological forecasts," in: Some Problems of Contemporary and Applied Hydrology [in Russian], part I, Izd-vo MGU, Moscow, 1981.
6. L. S. Kuchment and V. V. Puklakov, "Consideration of initial conditions and approximation of functions of the effect for short-range forecasts of water conditions," Trudy Gidromettsentra SSSR, no. 218, 1979.

10 May 1983

Hydrometeorological Scientific Research Center of the USSR